Next-Generation Sequencing (NGS) Analysis

This section provides a comprehensive introduction to NGS data analysis, covering essential computational skills, biological data handling, and practical workflows. Participants will gain hands-on experience using Linux-based environments and widely used bioinformatics tools for real-world genomic data analysis.

◇ Module 0 – Linux on Windows (WSL)

This module introduces the Linux environment, which is fundamental for bioinformatics workflows. Participants will understand why most bioinformatics tools are optimized for Linux systems and learn how to set up a Linux environment on Windows using Windows Subsystem for Linux (WSL).

- 0.1 Introduction to Linux and its importance in bioinformatics
- 0.2 Limitations of running bioinformatics tools on native Windows
- 0.3 Installation and setup of Linux using WSL
- 0.4 Basic Linux system orientation and navigation
- 0.5 Essential system preparation and command-line setup

◇ Module 1 – NGS Fundamentals & Linux Basics

This module covers the foundational concepts of NGS and introduces basic Linux commands required for handling sequencing data.

- 1.1 Introduction to Next-Generation Sequencing (NGS)
- 1.2 Overview of end-to-end NGS analysis workflow
- 1.3 Introduction to Linux shell environment
- 1.4 Hands-on practice with essential Linux commands
- 1.5 Understanding Linux file system structure
- 1.6 Setting up a mini-project workflow

◇ Module 2 – File Formats & Real NCBI Data

Participants will learn about biological data types, common file formats, and how to access and interpret real genomic data from public databases.

- 2.1 Types of biological data in genomics

- 2.2 Core file formats: FASTA, FASTQ, BAM, and VCF
- 2.3 Overview of public databases (NCBI, etc.)
- 2.4 Downloading reference genome (e.g., *E. coli*)
- 2.5 Understanding and interpreting FASTA file structure

◇ Module 3 – BLAST Analysis

This module introduces sequence similarity search using BLAST and compares local and web-based approaches.

- 3.1 Introduction to BLAST and its applications
- 3.2 Local BLAST vs. Web BLAST comparison
- 3.3 Hands-on execution of BLAST commands
- 3.4 Interpretation of BLAST output results
- 3.5 Comparative analysis using web-based BLAST tools

◇ Module 4 – Quality Control (QC)

Participants will learn how to assess and improve the quality of raw sequencing data using standard QC tools.

- 4.1 Importance of quality control in NGS
- 4.2 Understanding FASTQ format and Phred quality scores
- 4.3 Quality assessment of raw sequencing data
- 4.4 Read trimming and cleaning techniques
- 4.5 Post-processing quality evaluation
- 4.6 Interpretation of FastQC reports and modules

◇ Module 5 – Genome Assembly

This module focuses on assembling sequencing reads into complete genomes and evaluating assembly quality.

- 5.1 Concept of genome assembly
- 5.2 Applications and use-cases of assembly
- 5.3 Hands-on assembly using SPAdes
- 5.4 Validation and assessment of assembly results
- 5.5 Use of external tools for assembly confirmation

◇ Module 6 – Sequence Alignment

Participants will learn how sequencing reads are aligned to reference genomes and how alignment data is processed.

- 6.1 Introduction to sequence alignment
- 6.2 Workflow using BWA and SAMtools
- 6.3 Understanding SAM and BAM file formats
- 6.4 Evaluation of alignment metrics
- 6.5 Overview of alternative alignment tools

◇ Module 7 – Variant Calling (Conceptual)

This module introduces the principles of identifying genetic variations from sequencing data.

- 7.1 Definition and types of genetic variants
- 7.2 Basic logic of variant calling pipelines
- 7.3 Structure and interpretation of VCF files
- 7.4 Overview of industry-standard variant calling pipelines

Outcome of Section

By the end of this section, participants will be able to:

- Work efficiently in a Linux-based bioinformatics environment
- Handle real-world NGS datasets
- Perform key steps such as QC, assembly, alignment, and basic analysis
- Understand the complete NGS workflow from raw data to biological insights